A Statistical Review of the B.C. Forest Inventory Audit Sampling Scheme and Estimation Methods

Vera Sit and Peter Ott
Biometricians
B.C. Forest Service
Research Branch
Victoria, B.C.
V8W 9C2

April 1999

Background

In November 1998, the Chief Forester requested that the BC Forest Service Research Branch conduct a review of the application of forest inventory audit information in timber supply analysis supporting allowable annual cut determination. The review raised a number of statistical concerns regarding the audit sampling design and the approaches for adjusting the inventory data with the audit data (Research Branch, 1999). As a result, the Director of Resources Inventory Branch requested the Research Branch biometricians to perform further analyses to assess the current methods for data adjustment. This report presents findings of these latest analyses.

Inventory audit sampling scheme

The inventory audit uses an ordered systematic (OS) sampling scheme: an ordered list of all polygons in an inventory unit is first assembled; audit polygons are then systematically selected from the list at fixed intervals with a random starting point (Resources Inventory Branch, 1996). The sampling interval is determined by the total area of the inventory unit divided by the number of polygons to be sampled, and the polygons are selected with probability roughly proportional to polygon area. A cluster of plots is established within each selected polygon systematically using a square grid. A similar OS scheme is used in the current Vegetation Resources Inventory (VRI), except that an audit cluster attempts to cover an entire polygon uniformly, while a VRI cluster is restricted to within 50 m of a selected grid point.

Contrary to the common claim, the audit sampling scheme is **not** simple random sampling (SRS). A true SRS scheme would allow the occurrence of all possible samples with equal probability. For example, it would be possible to select the first 50 polygons in the ordered list in a true SRS scheme; such sample is impossible in the audit OS scheme. Although in theory, the OS sampling scheme has the property that every grid point in an inventory unit has an equal chance of being selected provided that there is a one-to-one correspondence between the area of a polygon and the number of grid points, this one-to-one correspondence is probably not true for large or irregular polygons (Resources Inventory Branch, 1996). See Taylor et al. (1998) for a more in-depth discussion on this point.

Furthermore, under the OS scheme, the probability that any two polygons have been included in a sample is not the same for all pairs of polygons — for some pairs of polygons, this probability is zero. This probability must be constant under a true SRS scheme (Thompson, 1992; page 22).

Presently, all adjustment procedures are based on SRS estimators. We investigated the appropriateness of using SRS estimators on audit data obtained by an OS sampling scheme.

Analysis objectives

The two main objectives are:

1. To assess the suitability of using OS, SRS, and probability proportional to size with replacement (PPSWR) estimators for adjusting inventory data by audit data obtained with an OS scheme.
2. Based on the findings in (1), to comment on the effectiveness of the current VRI sampling scheme.

Analysis procedure

Analyses were performed on audit data from three inventory units: the Fraser TSA, the Golden TSA, and the Queen Charlotte Island (QCI) TSA. For each set of audit data, we did the following calculations using OS, SRS, and PPSWR formulae. See Appendix A for estimation formulae.

1. Estimated the difference between audit volume and inventory volume (expressed in volume per hectare for the entire TSA) and the corresponding variance.
2. Estimated the age, top height, and volume adjustment ratios and the corresponding variances of the adjusted age, height, and volume (per hectare). Age and height ratios are used to adjust inventory age and height; volume ratios are used to adjust the revised inventory volume based on the adjusted age and height.

The Fraser TSA was post-stratified by inventory type group; QCI by inventory type group, site productivity, and age class; and Golden by age class. Calculations for (1) and (2) were done separately for each post-stratified stratum and for the entire inventory unit.

In addition, simulation analyses were performed to evaluate the relative bias of the OS, SRS, and PPSWR estimators of volume per hectare mean and variance, and age adjustment ratio for data sampled using an OS scheme. Similar relative bias computations were also done for PPSWR estimators on data obtained from PPSWR sampling scheme, and SRS estimators on data obtained from SRS sampling.

For the simulation study assessing mean volume per hectare and associated variance, a hypothetical population of 1200 polygons was generated using a normal distribution (with mean=626 and variance=56000) to assign volume per ha, and polygon areas were generated using an exponential distribution with mean area of 25 ha. Simulated values of volume per ha less than 150 m^3 per ha were truncated to 150. These distributions were chosen to roughly mimic the QCI audit data. One thousand samples of 50 polygons were independently chosen from the unsorted simulated data using the OS scheme and PPSWR scheme.

For the simulation studies assessing age adjustment ratio, two populations of 1200 polygons were generated to measure relative bias for a random population and a population with a periodic pattern. For the random population, the inventory polygons ages were normally distributed inventory with a mean of 250 years and variance 10000. For the patterned population, the inventory age of each polygon (inv_i) was generated using the following cosine model,

$$inv_i = A\cos\left(B \cdot 2\pi \cdot \frac{cum_i}{Z}\right) + A + \delta_i$$

where $A = 250$, $B = 50$, $cum_i = \sum_{j=1}^{i} area_j$, Z = total area of polygons in the population, and $\delta_i \sim N(0,1000)$. For both populations, polygon areas were generated using an exponential distribution (with mean 25 ha), and the audit age (aud_i) was modelled as

$$aud_i = 1.5 \cdot inv_i + \varepsilon_i$$

where $\varepsilon_i \sim N(0,\sigma^2)$. We repeated the simulations for population variances (σ^2) of 1000, 10000, and 100000 to check how the "tightness" of the relationship between audit and inventory age affect relative bias. One thousand samples of 50 polygons were

independently chosen using the OS scheme and PPSWR scheme. See Appendix B for equations used in the simulation analyses.

Results and discussion

Audit data analysis

The audit data analysis results for the Fraser, QCI, and Golden TSAs are shown in Tables 1, 2, and 3, respectively. Please refer to these tables in the upcoming section.

Age and height adjustment ratios: OS and PPSWR estimates are the same algebraically, but SRS estimates are different from those of OS and PPSWR, with the greatest difference about 20% (Fraser TSA, age adjustment ratio for C leading species stratum).

Volume adjustment ratios: OS, PPSWR, and SRS estimates are the same because of cancellation of the polygon area terms in the OS and PPSWR formulae, leaving a simplified equation identical to the SRS estimator equation. However, SRS estimates would not be the same as the OS and PPSWR estimates if the inventory unit contained polygons with selection probability equal to one (these polygons have areas greater than the selection interval). See Appendix C for the proof. Although the 3 TSAs in this study do not contain such large polygons, the Lilloet and Dawson Creek TSAs did have many polygons with area greater than the selection interval (Taylor et al. 1998).

Difference between audit volume and inventory volume: Stratified OS, PPSWR, and SRS estimates are identical for the same reasons given for the volume ratio results. See Appendix D for the proof. However, the stratified overall means are quite different from the stratum means; hence decisions based on the overall mean would not necessary apply to all strata. Moreover, the overall means computed with and without stratification can be very different (Table 3d). At present, adjustment decisions are based on the non-stratified overall mean but adjustments are made by stratum. The adjustment decision and the adjustment procedure should be made at the stratum level. If desired, use the stratified overall mean to describe the difference between audit and inventory volume for the entire TSA.

Variance: PPSWR and SRS variances are exactly equivalent for volume adjustment ratio (see Appendix E for proof) and approximately equal for differences in volume per ha (the

small discrepancy is due the presence of a finite population correction in the SRS equation). None of the OS variances were estimable for three reasons:

1. OS variances require the calculation of the joint selection probabilities of all pairs of polygons in the audit sample. These probabilities depend on the ordered list used for sample selection and the sampling interval. The OS sampling protocol defines the sampling interval to be the total area in the inventory unit divided by the number of polygons required in the sample. However, in practice, this protocol is not strictly followed. An excess number of polygons (about 500) are selected using a much smaller sampling interval; the chosen polygons are then randomly ordered and the top 50 polygons in this random list form the audit sample. Consequently, the resulting audit samples no longer follow an OS scheme, have unknown statistical properties, and the variances become unknown.
2. For the case of the Fraser TSA, two samples were taken at different times, using different sampling intervals. Combined (true) OS variances were not estimable in this case even if the OS sampling protocol had been followed.
3. Because of the nature of the OS sampling scheme, for many pairs of the sampled polygons (i, j), the probability of selecting polygon i is smaller than the conditional probability of selecting polygon j given polygon i has been selected. Consequently, OS variance estimates are often negative.

Because the OS variances are not estimable, OS estimators are not very useful for fully analysing the audit data — it is impossible to construct reliable tests of hypotheses or confidence intervals for any of the OS estimators. As a result, simulation analyses were performed to determine which of the two estimators, PPSWR or SRS, would be more suitable for approximating OS estimators under the OS sampling scheme.

Simulation analysis

Volume per hectare mean and variance: For mean volume per hectare estimation, on data obtained with an OS scheme, all 3 estimators (OS, SRS, and PPSWR) have the same relative bias; although PPSWR estimator on PPSWR data has slightly smaller relative bias, the bias is too small (less than one-tenth of a percent) to be of any concern (Table 4a). An SRS estimator based on a set of SRS samples is also presented as a numerical 'benchmark'. It indicates that a relative bias as high as 1.24% can be attributed to numerical imprecision (e.g. low number of sample realisations relative to the sampling

fraction) since a perfectly unbiased result is expected. For volume per hectare variance estimation, the OS estimator on OS data has the greatest relative bias due to negative estimated variances in all 1000 simulations. The relative bias of both SRS and PPSWR estimators is large with the SRS bias being slightly smaller. On the other hand, the PPSWR estimator from PPSWR samples has a slightly larger true variance but much smaller relative bias (Table 4b), which is expected since the estimator matches the sampling scheme.

The simulation results suggest that for volume per hectare estimation on audit data, both SRS and PPSWR mean estimators perform equally well as approximations for OS estimators but the relative bias in the variance is alarmingly high. Since a PPSWR sample with PPSWR estimators for mean volume per hectare and associated variance are unbiased (Thompson 1992) and should have small true variance, **a PPSWR sampling scheme with PPSWR estimators would provide more reliable results**.

Age ratio: For a randomly ordered polygon list, OS data with SRS estimator have the smallest relative bias, followed by OS data with PPSWR or OS estimator, and PPSWR data with PPSWR estimator (Table 5a). This pattern is observed in all cases of σ^2, with the discrepancy in relative bias among the estimators magnified for large σ^2. Because the ages in the polygon list are in random order, SRS estimators are acceptable for systematic samples — the relative bias for the SRS estimator is essentially zero even for the most variable case ($\sigma^2 = 100000$).

The same ranking of relative bias is entirely reversed for the population of polygons with periodic pattern (Table 5b). The magnitude of the bias increases drastically for all the estimators on OS samples; these biases become unacceptably large for moderate to large σ^2. On the other hand, the PPSWR estimator on PPSWR samples has negligible bias (under 2%) as in the random population. These results also would apply to height adjustment ratios.

Note that the relative biases from our simulation study represent long run results — the average bias over many samples. The relative bias based on a single sample could be much larger or smaller than those given by our simulations. The large differences between SRS and OS age adjustment ratios in the audit data (Tables 1a, 2a, and 3a) could be due to a non-random ordered list, or an unusual sample, or both. Be cautious that volume adjustment ratios are based on the revised inventory volume with age and height

adjustment ratios, and biased age and height adjustment ratios would propagate into the volume adjustment ratios.

Conclusions

The decision of whether to adjust the inventory and the adjustment procedure must be consistent. We recommend that both the adjustment decision and the adjustment process be made for individual strata. If desired, use the stratified overall mean to describe the difference between audit and inventory volume for the entire TSA.

Due to the nature of the OS sampling scheme, OS variance estimates are often negative. Since audit samples are selected without strict adherence to the OS design protocol, OS variance estimates based on the current audit data cannot be calculated. Hence it is impossible to analyse the audit data properly using OS estimators.

Both SRS and PPSWR estimators can be used to approximate OS estimators when analysing audit data. Although PPSWR estimators tend to have larger relative bias compared to SRS estimators, the differences are quite small in most cases; exception for age ratio with an underlying pattern in the ordered list. Therefore, for the purpose of adjusting the inventory data based on audit data, either SRS or PPSWR estimators are suitable under the assumption that:

1. The audit samples were obtained using an OS sampling scheme.
2. For any inventory unit, the audit samples came from a single selection.
3. There was no underlying pattern in the ordered list with respect to the adjustment variables (such as volume per hectare, age, and height).

The simulation study indicated that the SRS and PPSWR estimators tend to overestimate the variance in mean volume per hectare by about 30%, which would result in wider confidence intervals and less frequent conclusions of a significant difference. However, since assumption (1) is violated in the audit, and assumption (2) was not the case for some inventory units (e.g., Fraser TSA), the actual bias in the SRS and PPSWR estimators is unknown.

Volume adjustment ratios are based on revised inventory volumes with the adjusted ages and heights. Therefore, error in the age and height adjustment ratios will propagate to the volume adjustment ratios, inflating the errors in the final adjusted volume.

Knowing the problems associated with the OS sampling scheme, we strongly recommend changing the VRI sampling scheme from OS to PPSWR (possibly in stratified form) and using PPSWR estimators for all analyses. In other words, instead of selecting the polygons systematically, polygons would be chosen independently with probability proportional to their area. The change would only affect the polygon selection stage of the inventory, which usually happens in the office rather than in the field. If necessary, additional polygons can be chosen at a later time without affecting the statistical properties of a PPSWR sample. Perhaps another possible approach is to take several (at least three) independent OS samples from the same population and use the resulting set of means to calculate unbiased estimators of the overall mean and its associated variance. Either way, to realise all of the benefits of any sampling scheme, the sampling protocol must be followed precisely.

References

Cochran, W. G. 1977. Sampling Techniques, Third Edition. John Wiley & Sons Inc., NY.

Research Branch. 1999. Review of the application of forest inventory audit information in timber supply review and timber supply analyses. BC Min. Forests, Victoria, BC.

Resources Inventory Branch. 1996. Fraser TSA inventory audit. BC Min. Forests, Victoria, BC.

Taylor, C. and C. Schwarz. 1998. A review of the BC vegetation resources inventory sampling design and the proposed estimators. Prepared for the Resources Inventory Branch.

Thompson, S. K. 1992. Sampling. John Wiley & Sons Inc., NY.

Table 1a. Fraser TSA age adjustment ratios and variances of the adjusted age

		Ratios			Variances		
ITG	n	OS	PPSWR	SRS	OS	PPSWR	SRS
F, FPI, Fdecid.	31	0.73	0.73	0.87	?	1298.17	93.34
FC, FH, FB, FS, FPy	22	0.94	0.94	0.86	?	397.29	85.95
All C leading	5	1.00	1.00	1.21	?	4873.70	5884.70
All H or B leading	50	0.98	0.98	1.02	?	139.11	99.66
S or PI leading	7	0.98	0.98	0.95	?	630.49	665.35
All Decid. leading	9	0.90	0.90	0.76	?	22.80	36.12
Overall (not stratified)	124	0.88	0.88	0.96	?	269.70	38.81

Table 1b. Fraser TSA height adjustment ratios and variances of the adjusted height

		Ratios			Variances		
ITG	n	OS	PPSWR	SRS	OS	PPSWR	SRS
F, FPI, Fdecid.	31	1.00	1.00	1.02	?	25.14	1.46
FC, FH, FB, FS, FPy	22	0.82	0.82	0.88	?	20.99	2.01
All C leading	5	1.00	1.00	0.98	?	0.67	2.24
All H or B leading	50	0.81	0.81	0.90	?	1.91	0.84
S or PI leading	7	0.96	0.96	1.03	?	4.17	4.93
All Decid. leading	9	0.95	0.95	0.90	?	2.28	5.56
Overall (not stratified)	124	0.90	0.90	0.94	?	3.80	0.36

Table 1c. Fraser TSA volume adjustment ratios and variances of the adjusted volume

		Volume Ratios			Variances		
ITG	n	OS	PPSWR	SRS	OS	PPSWR	SRS
F, FPI, Fdecid.	31	0.77	0.77	0.77	?	406.14	406.14
FC, FH, FB, FS, FPy	22	0.89	0.89	0.89	?	2290.56	2290.57
All C leading	5	0.92	0.92	0.92	?	1196.52	1196.52
All H or B leading	50	0.86	0.86	0.86	?	836.48	836.40
S or PI leading	7	0.91	0.91	0.91	?	8201.54	8201.54
All Decid. leading	9	1.33	1.33	1.33	?	5309.89	5309.89
Overall (not stratified)	124	0.85	0.86	0.86	?	294.02	294.02

Table 1d. Fraser TSA estimated overall difference (audit volume - inventory volume) in volume per hectare and variance

		Mean (audit - inventory) volume per ha			Variances		
ITG	n	OS	PPSWR	SRS	OS	PPSWR	SRS
F, FPI, Fdecid.	31	-97.24	-97.24	-97.24	?	622.88	614.69
FC, FH, FB, FS, FPy	22	-152.06	-152.06	-152.06	?	2594.65	2574.99
All C leading	5	-158.02	-158.02	-158.02	?	1053.17	1047.95
All H or B leading	50	-180.98	-180.98	-180.98	?	1018.68	1009.45
S or PI leading	7	-14.47	-14.47	-14.47	?	9638.50	9518.87
All Decid. leading	9	25.87	25.87	25.87	?	5439.60	5354.90
Overall (not stratified)	124	-129.58	-129.58	-129.58	?	365.86	362.45
Overall (stratified)	124	-140.16	-140.16	-140.16	?	14.73	14.59

Table 2a. QCI TSA age adjustment ratios and variances of the adjusted age

		Age Ratios			Variances		
Species Class	**n**	**OS**	**PPSWR**	**SRS**	**OS**	**PPSWR**	**SRS**
Cedar-old	18	0.87	0.87	0.78	?	990.40	412.56
Cedar-young	11	0.99	0.99	1.09	?	52.83	234.00
Hemlock-good	14	0.98	0.98	0.98	?	398.46	333.17
Hemlock-poor	11	0.99	0.99	0.83	?	1405.75	1069.45
Spruce	11	0.78	0.78	0.80	?	166.35	688.97
Overall (not stratified)	65	0.90	0.90	0.86	?	166.04	107.99

Table 2b. QCI TSA height adjustment ratios and variances of the adjusted height

		Height Ratios			Variances		
Species Class	**n**	**OS**	**PPSWR**	**SRS**	**OS**	**PPSWR**	**SRS**
Cedar-old	18	0.96	0.96	0.90	?	6.38	1.04
Cedar-young	11	1.09	1.09	0.98	?	6.64	2.21
Hemlock-good	14	0.97	0.97	0.98	?	3.25	1.86
Hemlock-poor	11	1.05	1.05	1.03	?	1.50	2.01
Spruce	11	0.85	0.85	0.86	?	0.47	3.81
Overall (not stratified)	65	0.96	0.96	0.94	?	1.04	0.44

Table 2c. QCI TSA volume adjustment ratios and variances of the adjusted volume

		Volume Ratios			Variances		
Species Class	**n**	**OS**	**PPSWR**	**SRS**	**OS**	**PPSWR**	**SRS**
Cedar-old	18	1.31	1.31	1.31	?	2702.24	2702.24
Cedar-young	11	1.57	1.57	1.57	?	4349.78	4349.78
Hemlock-good	14	1.08	1.08	1.08	?	3403.29	3403.29
Hemlock-poor	12	1.04	1.04	1.04	?	4286.68	4286.68
Spruce	12	0.97	0.97	0.97	?	4927.4	4927.47
Overall (not stratified)	67	1.14	1.14	1.14	?	822.19	822.19

Table 2d. QCI TSA estimated overall difference
(audit volume - inventory volume) in volume per hectare and variance

		Mean (audit - inventory) volume per ha			Variances		
Species Class	**n**	**OS**	**PPSWR**	**SRS**	**OS**	**PPSWR**	**SRS**
Cedar-old	18	78.52	78.52	78.52	?	2495.15	2421.64
Cedar-young	11	195.27	195.27	195.27	?	3773.60	3609.53
Hemlock-good	14	41.42	41.42	41.42	?	3320.63	3188.93
Hemlock-poor	12	58.87	58.87	58.87	?	4322.97	4242.55
Spruce	12	-152.06	-152.06	-152.06	?	4753.77	4586.97
Overall (not stratified)	67	45.12	45.12	45.12	?	840.85	815.47
Overall (stratified)	67	44.13	44.13	44.13	?	59.13	57.54

Table 3a. Golden TSA age adjustment ratios and variances of the adjusted age

		Age Ratios			Variances		
Age Class	n	OS	PPSWR	SRS	OS	PPSWR	SRS
0 to 7	23	0.95	0.95	1.05	?	241.80	31.10
8 & 9	27	0.76	0.76	0.80	?	185.54	215.70
Overall (not stratified)	50	0.80	0.80	0.85	?	282.76	77.15

Table 3b. Golden TSA height adjustment ratios and variances of the adjusted height

		Height Ratios			Variances		
Age Class	n	OS	PPSWR	SRS	OS	PPSWR	SRS
0 to 7	23	1.03	1.03	1.07	?	6.79	1.39
8 & 9	27	0.92	0.92	0.93	?	2.52	0.81
Overall (not stratified)	50	0.96	0.96	0.99	?	4.89	0.58

Table 3c. Golden TSA volume adjustment ratios and variances of the adjusted volume

		Volume Ratios			Variances		
Age Class	n	OS	PPSWR	SRS	OS	PPSWR	SRS
0 to 7	23	1.34	1.34	1.34	?	780.65	780.48
8 & 9	27	0.98	0.98	0.98	?	892.89	893.64
Overall (not stratified)	50	1.12	1.12	1.12	?	509.14	509.14

Table 3d. Golden TSA estimated overall difference
(audit volume - inventory volume) in volume per hectare and variance

		Mean (audit - inventory) volume per ha			Variances		
Age Class	n	OS	PPSWR	SRS	OS	PPSWR	SRS
0 to 7	23	86.69	86.69	86.69	?	632.68	632.09
8 & 9	27	-6.38	-6.38	-6.38	?	917.58	915.72
Overall (not stratified)	50	36.43	36.43	36.43	?	437.78	436.89
Overall (stratified)	50	70.52	70.52	70.52	?	15.39	15.37

Simulation results to estimate relative bias in volume per ha mean and variance

Table 4a. Simulation results for mean volume per ha estimates. The data was not sorted for the implementation of the OS method.

Method	Estimator	True mean	Estimated mean	Relative Bias (%)
OS	SRS	616.93	617.43	0.08
OS	PPSWR	616.93	617.43	0.08
OS	OS	616.93	617.43	0.08
PPSWR	PPSWR	616.93	617.06	0.02
SRS	SRS	616.93	624.60	1.24

Table 4b. Simulation results for mean volume per ha variance estimates The data was not sorted for the implementation of the OS method.

Method	Estimator	True variance	Estimated variance	Relative Bias (%)
OS	SRS	882.01	1133.62	28.53
OS	PPSWR	882.01	1182.91	34.11
OS	OS	882.01	-47418.43	-5476.15
PPSWR	PPSWR	1155.73	1167.80	1.04
SRS	SRS	1110.74	1130.13	1.75

For the OS variance, estimates for all 1000 simulations are negative.

Table 5a. Simulation results for age adjustment ratios for a random population. The data was not sorted for the implementation of the OS method.

Method	Estimator	σ^2	Relative bias (%)
OS	SRS	1000	-0.04
OS	PPSWR or OS	1000	0.20
PPSWR	PPSWR	1000	0.34
OS	SRS	10000	-0.12
OS	PPSWR or OS	10000	0.60
PPSWR	PPSWR	10000	1.05
OS	SRS	100000	-0.34
OS	PPSWR or OS	100000	1.82
PPSWR	PPSWR	100000	3.19

Table 5b. Simulation results for age adjustment ratios for a population with periodic pattern. The data was not sorted for the implementation of the OS method (i.e. a periodic pattern remained in the population).

Method	Estimator	σ^2	Relative bias (%)
OS	SRS	1000	5.61
OS	PPSWR or OS	1000	4.46
PPSWR	PPSWR	1000	0.08
OS	SRS	10000	21.26
OS	PPSWR or OS	10000	18.32
PPSWR	PPSWR	10000	0.47
OS	SRS	100000	74.20
OS	PPSWR or OS	100000	67.08
PPSWR	PPSWR	100000	1.75

Notation used in the Appendices

SRS - simple random sampling (without replacement)
OS - ordered systematic sampling
PPSWR - probability proportional to size with replacement
str - stratified sampling
gr - generalized ratio
$\hat{r}$ - estimated ratio
z_i - area of polygon i
Z - total area of all polygons in population
Z_h - total area of polygons in stratum h
y_i - audit volume per hectare for polygon i
x_i - (height-age) adjusted volume per hectare for polygon i
Y_i - total audit volume for polygon $i = y_i z_i$
X_i - total (height-age) adjusted volume for polygon $i = x_i z_i$
n - number of polygons sampled
n_h - number of polygons sampled from stratum h
N - total number of polygons in population
N_h - total number of polygons in stratum h
k - sampling interval = Z/n
k_h - sampling interval for stratum $h = Z_h/n_h$
L - total number of strata
$\hat{y}_{NAME}$ - estimate of mean volume per ha
$v(\bullet)$ - estimated variance of the estimate
$\hat{y}_s$ - value of estimator for simulation s
S - total number of simulations

Appendix A. Estimation formulae for OS, PPSWR and SRS estimators

The following estimators can be also found in textbooks such as Cochran (1977), Thompson (1992), or the report by Taylor and Schwarz (1998). Note that all of these are single stage estimators (i.e. within-polygon variability is not considered).

Most of what follows is presented in terms of the variable: volume per ha. Parameter estimates for the variables age and height are almost identical (with the exclusion of the 'per area' component in the formulae) and hence, are not discussed.

SRS

The SRS estimate of mean volume per ha is given by

$$\hat{y}_{SRS} = \frac{1}{n}\sum_{i=1}^{n} y_i .$$

The estimated variance of this estimate is

$$v\left(\hat{y}_{SRS}\right) = \sum_{i=1}^{n} \frac{\left(y_i - \hat{y}_{SRS}\right)^2}{n(n-1)} \frac{(N-n)}{N} .$$

Generalized ratio estimators can be obtained using

$$\hat{y}_{gr_{SRS}} = \hat{r}\frac{\sum_{i=1}^{N} X_i}{Z} \quad \text{where} \quad \hat{r} = \frac{\hat{y}_{SRS}}{\hat{x}} \quad \text{and} \quad \hat{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

$$v\left(\hat{y}_{gr_{SRS}}\right) = \sum_{i=1}^{n} \frac{\left(y_i - \hat{r}x_i\right)^2}{n(n-1)} \frac{(N-n)}{N} .$$

PPSWR

The PPSWR estimate of mean volume per ha is given by

$$\hat{y}_{PPSWR} = \frac{1}{nZ}\sum_{i=1}^{n} \frac{Y_i}{p_i} \quad \text{where} \quad p_i = \frac{z_i}{Z} .$$

The estimated variance of this estimate is

$$v\left(\hat{y}_{PPSWR}\right) = \frac{1}{Z^2}\sum_{i=1}^{n} \frac{\left(\frac{Y_i}{p_i} - Z\hat{y}_{PPSWR}\right)^2}{n(n-1)} .$$

Generalized ratio estimators can be obtained using

$$\hat{y}_{gr_{PPSWR}} = \hat{r}\frac{\sum_{i=1}^{N} X_i}{Z} \quad \text{where} \quad \hat{r} = \frac{Z\hat{y}_{PPSWR}}{\hat{X}} \quad \text{and} \quad \hat{X} = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i}{p_i} .$$

$$v\left(\hat{y}_{gr_{PPSWR}}\right) = \frac{1}{Z^2} \frac{\sum_{i=1}^{n}\left(\frac{e_i}{p_i} - \frac{1}{n}\sum_{i=1}^{n} \frac{e_i}{p_i}\right)^2}{n(n-1)} \quad \text{where} \quad e_i = Y_i - \hat{r}X_i .$$

OS

The (Horvitz-Thompson) OS estimate of mean volume per ha is given by

$$\hat{y}_{OS} = \frac{1}{Z}\sum_{i=1}^{n}\frac{Y_i}{\pi_i} \quad \text{where} \quad \pi_i = \begin{cases} \dfrac{z_i}{k} & \text{if } z_i < k \\ 1 & \text{if } z_i \geq k \end{cases}.$$

The estimated variance of this estimate is

$$v\left(\hat{y}_{OS}\right) = \frac{1}{Z^2}\sum_{i=1}^{n}\sum_{j>i}^{n}\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 \quad \text{where } \pi_{ij} = \pi_i \cdot \pi_{j|i}.$$

Here, $\pi_{j|i} = \dfrac{\textit{area in polygon i that allows j to get selected}}{z_i}$.

Generalized ratio estimators can be obtained using

$$\hat{y}_{gr_{OS}} = \hat{r}\frac{\sum_{i=1}^{N} X_i}{Z} \quad \text{where} \quad \hat{r} = \frac{Z\hat{y}_{OS}}{\hat{X}} \quad \text{and} \quad \hat{X} = \sum_{i=1}^{n}\frac{X_i}{\pi_i}.$$

$$v\left(\hat{y}_{gr_{OS}}\right) = \frac{1}{Z^2}\sum_{i=1}^{n}\sum_{j>i}^{n}\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j}\right)^2 \quad \text{where} \quad e_i = Y_i - \hat{r}X_i.$$

SRS using total volume

A different SRS estimator, not considered in this document, could be derived using total volume per polygon (instead of volume per ha). This SRS estimate of mean volume per ha is given by

$$\tilde{y}_{SRS} = \frac{N}{nZ}\sum_{i=1}^{n} Y_i.$$

The estimated variance of this estimate is

$$v\left(\tilde{y}_{SRS}\right) = \frac{N^2}{Z^2}\sum_{i=1}^{n}\frac{\left(Y_i - \frac{Z}{N}\tilde{y}_{SRS}\right)^2}{n(n-1)}\frac{(N-n)}{N}.$$

Generalized ratio estimators can be obtained using

$$\tilde{\bar{y}}_{gr_{SRS}} = \hat{r}\frac{\sum_{i=1}^{N} X_i}{Z} \quad \text{where} \quad \hat{r} = \frac{Z\tilde{\bar{y}}_{SRS}}{\hat{X}} \quad \text{and} \quad \hat{X} = \frac{N}{n}\sum_{i=1}^{n} X_i .$$

$$v\left(\tilde{\bar{y}}_{gr_{SRS}}\right) = \frac{1}{Z^2}\sum_{i=1}^{n}\frac{\left(Y_i - \hat{r}X_i\right)^2}{n(n-1)}\frac{(N-n)}{N}.$$

Stratified Sampling

Stratified sampling occurs when the population is first divided into homogeneous strata and units are selected from each stratum independently. Poststratification occurs when polygons are classified into the various strata after the (SRS) sample is taken. Parameter estimates for individual strata can be obtained by making the appropriate substitutions in the above formulae (i.e. N_h=N, Z_h=Z, etc.). Obviously these estimates are derived assuming that the a particular sampling scheme (e.g. OS) was applied within each stratum. Combined overall stratified estimators can be obtained using the (pre-stratified) equations:

$$\hat{\bar{y}}_{str} = \frac{1}{N}\sum_{h=1}^{L} N_h \hat{\bar{y}}_h \quad \text{and}$$

$$v\left(\hat{\bar{y}}_{str}\right) = \sum_{h=1}^{L}\left(\frac{N_h}{N}\right)^2\left(\frac{N_h - n_h}{N_h}\right)\frac{s_h^2}{n_h} \qquad \text{where} \qquad s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\left(y_{hi} - \hat{\bar{y}}_h\right)^2 \quad .$$

If that data has been poststratified, the estimate $\hat{\bar{y}}_{str}$ still holds. However, the value of $v\left(\hat{\bar{y}}_{str}\right)$ may become different (likely larger) since the stratum sizes become random variables (Thompson 1992).

Appendix B. Simulation method

The true population mean volume per ha is given by $y = \frac{\sum_{i=1}^{N} Y_i}{Z}$,

and the estimated mean is $\bar{\hat{y}} = \frac{\sum_{s=1}^{S} \hat{y}_s}{S}$.

Therefore, relative bias for the mean is calculated using $RB = 100 \cdot \left(\frac{\bar{\hat{y}} - y}{y} \right)$.

The true variance is calculated using the mean squared error: $MSE = \frac{\sum_{s=1}^{S} \left(\hat{y}_s - \bar{\hat{y}} \right)^2}{S-1}$.

The estimated variance is $\overline{v(\hat{y})} = \frac{\sum_{s=1}^{S} v_s(\hat{y})}{S}$,

and so the relative bias in the variance is given by $RB = 100 \cdot \left(\frac{\overline{v(\hat{y})} - MSE}{MSE} \right)$.

For the simulation study of age adjustment ratios, analogous calculations to these were used (i.e. we used 'y' to denote mean age rather than mean volume per ha) with one exception: the true mean age was calculated using an area weighted mean. In other words, using the above notation,

true mean age = 'y' (in above notation) = $\frac{\sum_{i=1}^{N} z_i \cdot age_i}{Z}$.

Appendix C. Proof of equivalence among OS, SRS and PPSWR estimators for volume adjustment ratio

$$\hat{r}_{ITS} = \frac{\sum_{i=1}^{n} \frac{Y_i}{\pi_i}}{\sum_{i=1}^{n} \frac{X_i}{\pi_i}} = \frac{\sum_{i=1}^{n} \frac{y_i z_i}{\pi_i}}{\sum_{i=1}^{n} \frac{x_i z_i}{\pi_i}} = \frac{\sum_{i=1}^{n} y_i z_i \frac{k}{z_i}}{\sum_{i=1}^{n} x_i z_i \frac{k}{z_i}} = \frac{\sum_{i=1}^{n} y_i k}{\sum_{i=1}^{n} x_i k} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$

$$= \hat{r}_{SRS}.$$

Similarly, replace $\pi_i = \frac{z_i}{k}$ with $p_i = \frac{z_i}{Z}$ to show $\hat{r}_{PPS} = \hat{r}_{SRS}$.

If one or more polygons in the population have $z_i \geq k$, then the proof showing $\hat{r}_{OS} = \hat{r}_{SRS}$ does not hold since:

$$\hat{r}_{OS} = \frac{\sum_{i=1}^{n} \frac{Y_i}{\pi_i}}{\sum_{i=1}^{n} \frac{X_i}{\pi_i}} = \frac{\sum_{i=1}^{n} \frac{y_i z_i}{\pi_i}}{\sum_{i=1}^{n} \frac{x_i z_i}{\pi_i}} = \frac{\sum_{i=1}^{n_{c}(z_i<k)} y_i z_i \frac{k}{z_i} + \sum_{i=1}^{n_{c}(z_i \geq k)} \frac{y_i z_i}{1}}{\sum_{i=1}^{n_{c}(z_i<k)} x_i z_i \frac{k}{z_i} + \sum_{i=1}^{n_{c}(z_i \geq k)} \frac{x_i z_i}{1}} = \frac{\sum_{i=1}^{n_{c}(z_i<k)} y_i k + \sum_{i=1}^{n_{c}(z_i \geq k)} y_i z_i}{\sum_{i=1}^{n_{c}(z_i<k)} x_i k + \sum_{i=1}^{n_{c}(z_i \geq k)} x_i z_i}$$

$$\neq \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}.$$

Appendix D. Proof of equivalence among OS, SRS and PPSWR estimators for mean volume per hectare

$$\hat{y}_{OS} = \frac{1}{Z}\sum_{i=1}^{n} \frac{Y_i}{\pi_i} = \frac{1}{Z}\sum_{i=1}^{n} \frac{y_i z_i}{\pi_i} = \frac{1}{Z}\sum_{i=1}^{n} y_i z_i \frac{k}{z_i} = \frac{1}{Z}\sum_{i=1}^{n} y_i k = \frac{1}{Z}\sum_{i=1}^{n} y_i \frac{Z}{n} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$= \hat{y}_{SRS}.$$

As in Appendix C, this proof does not hold if any polygons have $z_i \geq k$.

Similarly, $\hat{y}_{PPS} = \frac{1}{nZ}\sum_{i=1}^{n} \frac{Y_i}{p_i} = \frac{1}{nZ}\sum_{i=1}^{n} y_i z_i \frac{Z}{z_i} = \hat{y}_{SRS}$.

It follows that the OS and PPS estimators of mean difference in volume per hectare (which are based on $D_i = Y_i - X_i$) are equivalent and equal to the SRS estimator (which is calculated using $d_i = y_i - x_i$).

Appendix E. Proof of the equivalence of PPSWR and SRS estimators for the variance of the ratio estimate of mean volume per hectare

$$v\left(\hat{y}_{gr_{PPSWR}}\right) = \frac{1}{Z^2}\frac{\sum_{i=1}^{n}\left(\frac{e_i}{p_i} - \frac{1}{n}\sum_{i=1}^{n}\frac{e_i}{p_i}\right)^2}{n(n-1)} \quad \text{where } e_i = Y_i - \hat{r}X_i \text{ and } \hat{r} = \frac{Z\hat{y}_{PPSWR}}{\hat{X}}.$$

$$v\left(\hat{y}_{gr_{PPSWR}}\right) = \frac{1}{Z^2}\frac{\sum_{i=1}^{n}\left(\frac{e_i}{p_i} - \frac{1}{n}\sum_{i=1}^{n}Z\frac{y_i z_i - \hat{r}x_i z_i}{z_i}\right)^2}{n(n-1)}$$

$$= \frac{1}{Z^2}\frac{\sum_{i=1}^{n}\left(\frac{e_i}{p_i} - \frac{Z}{n}\sum_{i=1}^{n}\left(y_i - \hat{r}x_i\right)\right)^2}{n(n-1)}$$

$$= \frac{1}{Z^2}\frac{\sum_{i=1}^{n}\left(\frac{e_i}{p_i} - \frac{Z}{n}\sum_{i=1}^{n}\left(y_i - \left(\frac{\sum_{i=1}^{n}y_i}{\sum_{i=1}^{n}x_i}\right)x_i\right)\right)^2}{n(n-1)}$$

$$= \frac{1}{Z^2}\frac{\sum_{i=1}^{n}\left(\frac{e_i}{p_i} - 0\right)^2}{n(n-1)}$$

$$= \frac{1}{Z^2}\frac{\sum_{i=1}^{n}\left(Z\left(y_i - \hat{r}x_i\right)\right)^2}{n(n-1)}$$

$$= \frac{\sum_{i=1}^{n}\left(y_i - \hat{r}x_i\right)^2}{n(n-1)}$$

$$\equiv v\left(\hat{y}_{gr_{SRS}}\right)$$